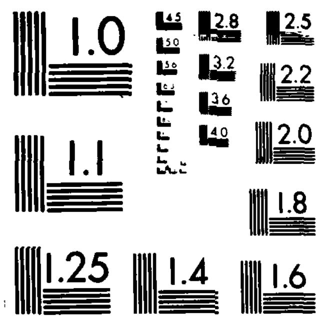
AD-A154 806 THE DISTRIBUTION OF THE NUMBER OF EMPTY CELLS IN A GENERALIZED RANDOM ALL..(U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER B HARRIS ET AL. MAR 85 1/1

UNCLASSIFIED MRC-TSR-2805 DAAG29-80-C-0041 F/G 12/1 NL

END
DATE
FILMED
7-85
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

MRC Technical Summary Report #2805

THE DISTRIBUTION OF THE NUMBER OF EMPTY CELLS IN A GENERALIZED RANDOM ALLOCATION SCHEME

Bernard Harris, Morris Marden
and C. J. Park

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

March 1985

(Received March 7, 1985)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

THE DISTRIBUTION OF THE NUMBER OF EMPTY CELLS IN A GENERALIZED RANDOM ALLOCATION SCHEME*

Bernard Harris, Morris Marden[1] and C. J. Park[2]

Technical Summary Report #2805
March 1985

ABSTRACT

The n balls are randomly distributed into N cells, so that no cell may contain more than one ball. This process is repeated m times. In addition, balls may disappear; such disappearances are independent and identically Bernoulli distributed. Conditions are given under which the number of empty cells has an asymptotically $(N \to \infty)$ standard normal distribution.

approaches infinity

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A/1	

AMS (MOS) Subject Classifications: 60C05, 60F05, 05A15

Key Words: Empty cells, Occupancy

Work Unit Number 4 (Statistics and Probability)

*Technical Report #759, Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706.

[1]University of Wisconsin-Milwaukee.

[2]San Diego State University.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

Some asymptotic properties of an occupancy model which includes many classical models as special cases are studied.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

THE DISTRIBUTION OF THE NUMBER OF EMPTY CELLS IN A GENERALIZED RANDOM ALLOCATION SCHEME*

Bernard Harris, Morris Marden[1] and C. J. Park[2]

1. INTRODUCTION

The distribution of the number of empty cells in the following random allocation process is considered. Let n, N be positive integers with $n \le N$. Assume that n balls are randomly distributed into N cells, so that no cell may contain more than one ball. Then, the probability that each of n specified cells will be occupied is $\binom{N}{n}^{-1}$. This process is repeated m times, so that there are $\binom{N}{n}^m$ random allocations of nm balls among the N cells. In addition, for each ball, let p, $0 \le p \le 1$, be the probability that the ball will not "disappear" from the cell. The "disappearances" are assumed to be stochastically independent for each ball; thus the disappearances constitute a sequence of nm Bernoulli trials.

Several special cases of this problem have previously been considered . In particular, $p = 1$, $n = 1$ is the classical occupancy problem, see [2],[3],[10]. The case $p = 1$, n arbitrary has been discussed in [4] and [7]. The case $0 < p < 1$, $n = 1$ is treated in C. J. Park [5].

In this paper, we obtain the probability distribution and moments of the number of empty cells. In section 3, we show that the number of

---

*Technical Report #759, Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706.

[1]University of Wisconsin-Milwaukee.

[2]San Diego State University.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

empty cells may be represented as a sum of independent Bernoulli random variables. This representation permits us to determine conditions on m, n, p, N such that the number of empty cells is asymptotically normally distributed.

This random allocation process may be viewed as a filing or storage process. Objects are randomly assigned to files or storage bins. From time to time, objects may be missing or have disappeared.

2. THE PROBABILITY DISTRIBUTION AND THE MOMENTS OF THE NUMBER OF EMPTY CELLS

Let m,n,N be positive integers with $n \leq N$. m sets, each consisting of n balls, are distributed into N cells at random so that no cell can contain more than one ball from the same set. As each set is distributed, the balls that have been placed during the preceding distributions are left in the cells. Thus, at the end of the process, cells may contain as many as m balls. In addition, each ball may "disappear" with common probability $1 - p$, $0 \leq p \leq 1$. These disappearances are stochastically independent and thus constitute a sequence of mn *Bernoulli trials*.

Let $P_{m,n,N,p}(j)$ be the probability that exactly j of the N cells are empty.

We now establish the following theorem.

__Theorem 1.__

$$P_{m,n,N,p}(j) = \binom{N}{n}^{-m}\binom{N}{j}\sum_{\ell=0}^{N-j}(-1)^{\ell}\binom{N-j}{\ell}\cdot$$

$$\left[\sum_{i=0}^{j+\ell}(1-p)^{i}\binom{N-j-\ell}{n-i}\binom{j+\ell}{i}\right]^{m}, \qquad 0 \leq j \leq N . \qquad (1)$$

__Proof.__ Let A_ν be the event that the νth cell is empty, $\nu = 1,2,\ldots,N$. Then,

$$P(A_\nu) = \binom{N}{n}^{-m}\left[\sum_{i=0}^{1}\binom{N-1}{n-i}(1-p)^i\right]^m . \tag{2}$$

For $1 \le \nu_1 < \nu_2 \le N$,

$$P(A_{\nu_1} \cap A_{\nu_2}) = \binom{N}{n}^{-m}\left[\sum_{i=0}^{2}\binom{N-2}{n-i}\binom{2}{i}(1-p)^i\right]^m . \tag{3}$$

Thus, for $1 \le \nu_1 < \nu_2 < \cdots < \nu_k \le N$,

$$P(A_{\nu_1} \cap A_{\nu_2} \cap \cdots \cap A_{\nu_k}) = \binom{N}{n}^{-m}\left[\sum_{i=0}^{k}\binom{N-k}{n-i}\binom{k}{i}(1-p)^i\right]^m . \tag{4}$$

Thus, using the inclusion-exclusion method, the probability that exactly j cells are empty is

$$P_{m,n,N,p}(j) = \binom{N}{n}^{-m}\sum_{r=j}^{N}\binom{N}{r}(-1)^{r-j}\binom{r}{j}\left[\sum_{i=0}^{r}\binom{N-r}{n-i}\binom{r}{i}(1-p)^i\right]^m . \tag{5}$$

We can write (5) in the form (1) by letting $r = j + \ell$.

We now determine the factorial moments of S, the number of empty cells.

<u>Theorem 2</u>. The νth factorial moment of S,

$$E(S^{(\nu)}) = \binom{N}{n}^{-m} N^{(\nu)} \left[\sum_{j=0}^{\nu}(1-p)^j\binom{N-\nu}{n-j}\binom{\nu}{j}\right]^m . \tag{6}$$

Proof. From J. Riordan [9], p. 53, from (4), it follows immediately that

$$E(S^{(\nu)}) = \binom{N}{\nu}\nu!\binom{N}{n}^{-m}\left[\sum_{i=0}^{\nu}\binom{N-\nu}{n-i}\binom{\nu}{i}(1-p)^{i}\right]^{m} . \tag{7}$$

We thus obtain the following.

Corollary. $$E(S) = N\left(1 - \frac{pn}{N}\right)^{m} , \tag{8}$$

$$\sigma_S^2 = N(N-1)\left[\frac{(N-n)(N-n-1)}{N(N-1)} + 2(1-p)\frac{n(N-n)}{N(N-1)} + (1-p)^2\frac{n(n-1)}{N(N-1)}\right]^{m}$$

$$+ N\left(1 - \frac{pn}{N}\right)^{m}\left[1 - N\left(1 - \frac{pn}{N}\right)^{m}\right] . \tag{9}$$

Proof. From (7)

$$E(S) = N\binom{N}{n}^{-m}\left(\binom{N-1}{n} + \binom{N-1}{n-1}(1-p)\right)^{m} = N\left(1 - \frac{pn}{N}\right)^{m} .$$

Since

$$\sigma_S^2 = E(S^{(2)}) + E(S) - (E(S))^2 ,$$

the conclusion follows readily from (6), after some elementary calculations.

For some purposes, the following equivalent forms of (9) will prove useful.

$$\sigma_S^2 = N(N-1)[1 - \frac{np(2(N-1)-p(n-1))}{N(N-1)}]^m + N(1-\frac{pn}{N})^m(1-N(1-\frac{pn}{N})^m) \quad (10)$$

and

$$\sigma_S^2 = N^2(1-\frac{pn}{N})^{2m}\{[1-\frac{np^2(N-n)}{(N-1)(N-pn)^2}]^m - 1\}$$

$$+ N(1-\frac{pn}{N})^m\{1 - (1-\frac{pn}{N})^m[1-\frac{np^2(N-n)}{(N-1)(N-pn)^2}]^m\}. \quad (11)$$

From Theorem 2, we readily obtain the following.

<u>Theorem 3</u>. The factorial moment generating function of S is given by

$$\phi_m(t) = E(1+t)^S = \sum_{r=0}^{N} \binom{N}{r} t^r \binom{N}{n}^{-m} \left(\sum_{j=0}^{r} (1-p)^j \binom{N-r}{n-j}\binom{r}{j}\right)^m . \quad (12)$$

Note that $\phi_m(t)$ is a polynomial in t of degree N. This fact is exploited in the next section, where the asymptotic distribution of S is obtained. In particular,

$$\phi_0(t) = (1+t)^N \quad (13)$$

and

$$\phi_1(t) = (1+t)^{N-n}(1 + (1-p)t)^n . \tag{14}$$

We now investigate the asymptotic distribution properties of the number of empty cells.

3. THE ASYMPTOTIC DISTRIBUTION OF THE NUMBER OF EMPTY CELLS

In this section, we determine conditions under which the number of empty cells (when suitably normalized) has an asymptotically normal distribution. In order to establish this, a number of preliminary results are required.

Lemma 1. Let N, n, r be non-negative integers, $r \le n \le N$. Then

$$\sum_{\nu=\alpha}^{r} \binom{r}{\nu}\binom{\nu}{\alpha}\binom{N-r}{n-\nu} = \binom{r}{\alpha}\binom{N-\alpha}{n-\alpha} . \tag{15}$$

Proof. Since $\binom{\nu}{\alpha} = 0$ whenever $\nu < \alpha$, we can write

$$\sum_{\nu=\alpha}^{r} \binom{r}{\nu}\binom{\nu}{\alpha}\binom{N-r}{n-\nu} = \sum_{\nu=0}^{r} \binom{r}{\nu}\binom{\nu}{\alpha}\binom{N-r}{n-\nu} .$$

To obtain the conclusion, note that

$$\sum_{x=0}^{r} \frac{\binom{x}{\alpha}\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = E\{X^{(\alpha)}\}/\alpha! ,$$

where X has the hypergeometric distribution. From B. Harris [1], p. 105,

$$\sum_{x=0}^{r} \frac{\binom{x}{\alpha}\binom{n}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{r^{(\alpha)} n^{(\alpha)}}{N^{(\alpha)} \alpha!} .$$

The conclusion follows immediately.

Lemma 2.

$$\sum_{j=0}^{n} \frac{(-1)^j \binom{n}{j}\binom{r}{j} p^j}{\binom{N}{j}} = \sum_{\nu=0}^{r} \frac{(1-p)^\nu \binom{N-r}{n-\nu}\binom{r}{\nu}}{\binom{N}{n}} \quad . \tag{16}$$

Proof. The right-hand side of (16) may be written

$$\sum_{\nu=0}^{r} \frac{\binom{N-r}{n-\nu}\binom{r}{\nu}}{\binom{N}{n}} \sum_{j=0}^{\nu} \binom{\nu}{j}(-1)^j p^j = \frac{\sum_{j=0}^{r} (-1)^j p^j \sum_{\nu=j}^{r} \binom{\nu}{j}\binom{N-r}{n-\nu}\binom{r}{\nu}}{\binom{N}{n}} \quad .$$

Thus, the coefficient of p^j is

$$(-1)^j \sum_{\nu=j}^{r} \binom{\nu}{j}\binom{N-r}{n-\nu}\binom{r}{\nu}/\binom{N}{n} \quad .$$

From Lemma 1,

$$(-1)^j \sum_{\nu=j}^{r} \binom{\nu}{j}\binom{N-r}{n-\nu}\binom{r}{\nu}/\binom{N}{n} = (-1)^j \binom{r}{j}\binom{N-j}{n-j}/\binom{N}{n} \quad ,$$

from which the conclusion follows immediately. Employing the above lemmas, we can now establish the following theorem.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2805	2. GOVT ACCESSION NO. AD-A154806	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE DISTRIBUTION OF THE NUMBER OF EMPTY CELLS IN A GENERALIZED RANDOM ALLOCATION SCHEME		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Bernard Harris, Morris Marden and C. J. Park		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office, P.O. Box 12211, Research Triangle Park, North Carolina 27709		12. REPORT DATE March 1985
		13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Empty cells
Occupancy

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

n balls are randomly distributed into N cells, so that no cell may contain more than one ball. This process is repeated m times. In addition, balls may disappear; such disappearances are independent and identically Bernoulli distributed. Conditions are given under which the number of empty cells has an asymptotically $(N \to \infty)$ standard normal distribution.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REFERENCES

1. Harris, B. (1966). Theory of Probability, Addison-Wesley Publishing Company, Reading, Mass.

2. Harris, B. and Park, C. J. (1971). "A note on the asymptotic normality of the distribution of the number of empty cells in occupancy problems," Ann. Inst. Statist. Math., 23, 507-513.

3. Holst, Lars (1977). "Some asymptotic results for occupancy problems," Ann. Probability, 5, 1028-1035.

4. Kolchin, V. F., Sevast'yanov, B. A. and Chistyakov, V. P. (1978). Random Allocations. V. H. Winstons & Sons, Washington, DC.

5. Marden, M. (1966). Geometry of Polynomials. Second Edition, Mathematical Surveys, No. 3, American Mathematical Society, Providence, R.I.

6. Park, C. J. (1972). "A note on the classical occupancy problem," Ann. Math. Statist., 43, 1698-1701.

7. Park, C. J. (1981). "On the distribution of the number of unobserved elements when m-samples of size n are drawn from a finite population population," Comm. Statist., A-Theory Methods, 10, 371-383.

8. Renyi, A. (1962). "Three new proofs and a generalization of a theorem of Irving Weiss," Magyar Tnd. Akad. Math. Kutató. Int. Közl. A, 7, 203-214.

9. Riordan, J. (1958). An Introduction to Combinatorial Analysis, John Wiley and Sons, Inc., New York, N.Y.

10. Sevast'yanov, B. A. and Chistyakov, V. P. (1964). "Asymptotic normality in the classical ball problem," Theory of Probability and Its Applications, 9, 198-211.

and

$$\frac{\kappa_2^{3/2}}{N} \to \infty \quad \text{whenever} \quad \frac{mnp}{N^{\left(1-\frac{1}{3(\rho+1)}\right)}} \to \infty \; .$$

The conclusion is obvious whenever $\frac{mnp}{N} \to r > 0$.

If $\alpha \to \infty$ as $N \to \infty$, then

$$\kappa_2 = Ne^{-\alpha} + O(Ne^{-2\alpha})$$

and

$$\frac{\kappa_2^{3/2}}{N} \to \infty \quad \text{whenever} \quad 3\alpha - \log N \to -\infty.$$

<u>Proof.</u> From (11), we can write, for $\alpha \to 0$,

$$\kappa_2 = N(e^{-\alpha})(1-e^{-\alpha} - \alpha p e^{-\alpha}) + O(np\alpha) + O(p^2\alpha^2)$$

where $\alpha = \frac{mnp}{N}$. Then, as $\alpha \to 0$,

$$\kappa_2 = N(1-\alpha+\alpha^2/2)(\alpha - \frac{\alpha^2}{2} - \alpha p + \alpha^2 p) + O(N\alpha^3) + O(mn\alpha).$$

Then, if $p \to p^* \neq 1$,

$$\kappa_2 = N\alpha(1-p) + O(N\alpha^2)$$

and

$$\frac{\kappa_2^{3/2}}{N} \to \infty \quad \text{whenever} \quad \frac{mnp}{N^{2/3}} \to \infty .$$

Similarly, if $(1-p) = c(\frac{mnp}{N})^\rho + O((\frac{mnp}{N})^\rho)$, $0 < \rho < 1$, $c > 0$, then

$$\kappa_2 = N\alpha(1-p) + o(N\alpha(1-p))$$

<u>Theorem 6</u>. $V = (S-E(S))/\sigma_S$ has an asymptotically standard normal distribution as $N \to \infty$, whenever any of the following conditions are satisfied.

1. $\frac{mpn}{N} \to 0, \quad p \to p^* \neq 1$ and $\frac{mnp}{N^{2/3}} \to \infty$;

2. $\frac{mpn}{N} \to 0, \quad (1-p) \to 0$ so that for some $c > 0$,

   $$(1-p) = c(\tfrac{mnp}{N})^{\rho} + o((\tfrac{mnp}{N})^{\rho}), \quad 0 < \rho < 1, \text{ and}$$

   $$\frac{mnp}{N^{\left(1-\frac{1}{3(\rho+1)}\right)}} \to \infty ;$$

3. $\frac{mnp}{N} \to 0, \quad (1-p) = c(\frac{mnp}{N})^{\rho} + o((\frac{mnp}{N})^{\rho}), \quad \rho \geq 1$, and

   $$\frac{mnp}{N^{5/6}} \to \infty ;$$

4. $\frac{mnp}{N} \to r > 0$;

5. $\frac{mnp}{N} \to \infty$ and $\frac{3mpn}{N} - \log N \to -\infty$.

From (29),

$$\log \phi_m(t) = nm \log(1-p) + \sum_{j=1}^{N} \log(t-t_j^{(m)}) = \sum_{j=1}^{N} \log(1+\tau_j t)$$

$$= \sum_{j=1}^{N} \sum_{k=1}^{\infty} \frac{(\tau_j t)^k}{k} (-1)^k .$$

Thus,

$$\frac{\kappa_{[\nu]}}{\nu!} = \sum_{j=1}^{N} \frac{(-1)^\nu}{\nu} \tau_j^\nu, \quad 0 < \tau_j \le 1 ,$$

and

$$|\kappa_{[\nu]}|/\nu! \le \frac{1}{\nu} \sum_{j=1}^{N} |\tau_j^\nu| \le N/\nu .$$

Then

$$|\sum_{j=1}^{\ell} \beta_{j,\ell} \kappa_{[j]}| \le c_\ell N, \tag{30}$$

since the $\beta_{j,\ell}$ do not depend on $N, n, m,$ or p .

We now establish the following theorem.

$$f(x) = c(x-x_1)(x-x_2)\cdots(x-x_N), \quad x_1 \le x_2 \le \cdots \le x_N ,$$

the representation follows by setting $\tau_j = -(t_j^{(m)})^{-1}$ and noting that $\xi(0) = \phi_m(0) = 1$.

Let $\kappa_\ell = \kappa_\ell(n,N,m,p)$ be the cumulants of S and let $\kappa_{[\nu]}$ be the factorial cumulants of S. That is,

$$\log \phi_m(t) = \sum_{\nu=1}^{\infty} \kappa_{[\nu]} t^\nu/\nu! \quad .$$

Then

$$\kappa_\ell = \sum_{j=1}^{\ell} \beta_{j,\ell} \kappa_{[j]}, \qquad \ell \ge 2,$$

where $\beta_{j,\ell}$ are the Stirling numbers of the second kind.

Then, as $N \to \infty$,

$$V = (S - E(S))/\sigma_S$$

is asymptotically distributed by the standard normal distribution (mean 0, variance unity), whenever

$$\kappa_\ell/\kappa_2^{\ell/2} \to 0 , \qquad \ell > 2 .$$

Proof. Let Y be a Bernoulli random variable with $P\{Y = 1\} = \tau$. Then the factorial moment generating function of Y is

$$E_Y\{(1+t)^Y\} = (1+\tau t).$$

If

$$W = \sum_{i=1}^{N} Y_j ,$$

where $Y_1, Y_2, \ldots, Y_N$ are mutually independent Bernoulli random variables with $P\{Y_j = 1\} = \tau_j$, then the factorial moment generating function of W is

$$\xi(t) = E_W\{(1+t)^W\} = \prod_{j=1}^{N} E_{Y_j}\{(1+t)^{Y_j}\} = \prod_{j=1}^{N} (1+\tau_j t), \qquad (28)$$

where $0 \le \tau_j \le 1$, $j = 1,2,\ldots,N$. From Theorem 4, the factorial moment generating fraction of S may be written

$$\phi_m(t) = (1-p)^{nm} \prod_{j=1}^{N} (t-t_j^{(m)}), \quad m = 0,1,\ldots, \qquad (29)$$

where $t_j^{(m)}$ are real and $-(1-p)^{-m} \le t_j^{(m)} \le -1$, $j = 1,2,\ldots,N$.

Since every polynomial of degree N with real roots has a unique representation of the form

The zeros of $\phi_0(t)$ are $t_1^{(0)} = t_2^{(0)} = \ldots = t_N^{(0)} = -1$. The zeros of $\phi_1(t)$ are $t_1^{(1)} = -1,\ldots,t_{N-n}^{(1)} = -1$, $t_{N-n+1}^{(1)} = -1/(1-p),\ldots,$ $t_N^{(1)} = -1/(1-p)$. Now apply Lemma 3 with $\psi(z) = \phi_1(z)$ obtaining $a = 1$, $b = (1-p)^{-1}$. Then, the zeros of $\phi_2(t)$ are real and satisfy

$$-(1-p)^{-2} \le t_j^{(2)} \le -1, \quad j = 1,2,\ldots,N .$$

It then follows readily by induction that the zeros of $\phi_k(t)$ are real and satisfy

$$-(1-p)^{-k} \le t_j^{(k)} \le -1, \quad j = 1,2,\ldots,N, \ k = 2,3,\ldots .$$

Theorem 5. For $1 \le n \le N$, $0 \le p \le 1$, $m \ge 1$, S has a representation as the sum of N mutually independent Bernoulli random variables. That is, there exist mutually independent Bernoulli random variables, $Y_j = Y_j(N,m,p,n)$, $j = 1,2,\ldots,N$, such that

$$S = \sum_{j=1}^{N} Y_j \tag{26}$$

and

$$P\{Y_j = 1\} = \gamma_j = 1 - P\{Y_j = 0\} . \tag{27}$$

$$B_p = \bigcap_{-\infty<\gamma<\infty} B_{p,\gamma} = \{z : z \text{ real}, -b(1-p)^{-1} \le x \le -a(1-p)^{-1}\}. \quad (25)$$

Consequently, $C \cup B_p$ is contained in the interval (21), proving the lemma.

We now establish the following theorem.

<u>Theorem 4</u>. Let

$$\phi_m(t) = \sum_{r=0}^{N} \binom{N}{r} t^r \binom{N}{n}^{-m} \left(\sum_{j=0}^{r} (1-p)^j \binom{N-r}{n-j} \binom{r}{j} \right)^m .$$

Let $t_1^{(m)}, t_2^{(m)}, \ldots, t_N^{(m)}$ be the zeros (not necessarily distinct) of $\phi_m(t)$. Then $t_j^{(m)}$, $j = 1,2,\ldots,N$ are all real and

$$-(1-p)^{-m} \le t_j^{(m)} \le -1, \quad j = 1,2,\ldots,N; m = 0,1,\ldots .$$

<u>Proof</u>. From (19),

$$\phi_{m+1}(t) = T(\phi_m(t)), \quad m = 0,1,\ldots,$$

and from (13),

$$\phi_0(t) = (1+t)^N .$$

That is, $\psi_1^*(z_1^{(1)},z_2^{(1)},\ldots,z_N^{(1)}) = T(\psi(z^*))$ is a linear symmetric function of $z_1^{(1)},z_2^{(1)},\ldots,z_N^{(1)})$. Thus, the conditions of Walsh's theorem (M. Marden [5], p. 62) are satisfied. Thus, if $z_1^{(0)},z_2^{(0)},\ldots,z_N^{(0)}$ are points in C_γ, then there is at least one point ζ in C_γ such that

$$T[(z^* - \zeta)^N] = 0 ,$$

that is, one can set $z_1^{(1)} = \zeta$, $z_2^{(1)} = \zeta,\ldots,z_N^{(1)} = \zeta$ and preserve the value 0. From (18),

$$T[(z^* - \zeta)^N] = (z^* - \zeta)^{N-n}(z^* - \zeta - pz^*)^n = 0 .$$

Thus either $z^* = \zeta$ and therefore z^* is in C_γ or $z^* = \zeta(1-p)^{-1}$ and z^* is in

$$B_{p,\gamma} = \{z:|z+(c-i\gamma)(1-p)^{-1}| \le [(c-a)^2 + \gamma^2]^{1/2}(1-p)^{-1} . \qquad (23)$$

However, γ is real and arbitrary. Hence it is clear that

$$C = \bigcap_{-\infty<\gamma<\infty} C_\gamma = \{z:z \text{ real}, -b \le x \le -a\} \qquad (24)$$

and

and

$$\psi_1(z) = T(\psi(z)) = c_1 \prod_{\alpha=1}^{N} (z-z_\alpha^{(1)}). \tag{20}$$

If the zeros of $\psi(z)$ are real and satisfy

$$-b \le x_\alpha \le -a, \quad a,b \ge 0 ,$$

then the zeros of $\psi_1(z)$ are real and satisfy

$$-\frac{b}{(1-p)} \le x_\alpha^{(1)} \le -a . \tag{21}$$

<u>Proof</u>. Let

$$C_\gamma = \{z: |z+(c-i\gamma)| \le [(c-a)^2+\gamma^2]^{1/2},\ c = \tfrac{1}{2}(a+b)\}. \tag{22}$$

Clearly $-a$ and $-b$ are on the boundary of the circular region C_γ. Consequently all zeros of $\psi(z)$ are in C_γ. Let z^* be a zero of $\psi(z)$. Let

$$\psi_1^*(z_1^{(1)},z_2^{(1)},\ldots,z_N^{(1)}) = c_1(z^*-z_1^{(1)})(z^*-z_2^{(1)}) \cdots (z^*-z_N^{(1)}).$$

$$\left(\sum_{j=0}^{n} \frac{(-1)^j \binom{n}{j} (pt)^j}{N^{(j)}} D^j\right) \sum_{r=0}^{N} \frac{\binom{N}{r} t^r}{\binom{N}{n}^k} \left(\sum_{\alpha=0}^{r} (1-p)^{\alpha} \binom{N-r}{n-\alpha} \binom{r}{\alpha}\right)^k$$

$$= \sum_{j=0}^{n} \frac{(-1)^j \binom{n}{j} (pt)^j}{N^{(j)}} \sum_{r=0}^{N} \binom{N}{r} r^{(j)} t^{r-j} \left(\sum_{\alpha=0}^{r} \frac{(1-p)^{\alpha} \binom{N-r}{n-\alpha} \binom{r}{\alpha}}{\binom{N}{n}}\right)^k ,$$

$$= \sum_{r=0}^{N} t^r \sum_{j=0}^{n} \frac{(-1)^j \binom{n}{j} \binom{r}{j}}{\binom{N}{j}} p^j \left(\sum_{\alpha=0}^{r} \frac{(1-p)^{\alpha} \binom{N-r}{n-\alpha} \binom{r}{\alpha}}{\binom{N}{n}}\right)^k .$$

The conclusion now follows from Lemma 2.

Let

$$T(f(t)) = \left(\sum_{j=0}^{n} \frac{(-1)^j \binom{n}{j} (pt)^j}{N^{(j)}} D^j\right) f(t), \quad 0 < p < 1 . \tag{18}$$

Then, from Theorem 3, we have that

$$\phi_{m+1}(t) = T(\phi_m(t)), \quad \phi_0(t) = (1+t)^N . \tag{19}$$

<u>Lemma 3</u>. Extend the domain of T to the complex plane, letting $z = x + iy; x,y$ real. Let

$$\psi(z) = \prod_{\alpha=1}^{N} (z-z_{\alpha})$$

Theorem 3. The factorial moment generating function of the number of empty cells $\phi_m(t)$ (12) satisfies the following differential-difference equation,

$$\phi_{m+1}(t) = \left(\sum_{j=0}^{n} \frac{(-1)^j \binom{n}{j} (pt)^j D^j}{N^{(j)}} \right) \phi_m(t), \quad m = 0,1,\ldots, \tag{17}$$

where $D^j = \dfrac{d^j}{dt^j}$.

Proof. For $m = 0$, $\phi_0(t) = (1+t)^N$; hence

$$\left(\sum_{j=0}^{n} \frac{(-1)^j \binom{n}{j} (pt)^j D^j}{N^{(j)}} \right) (1+t)^N = \sum_{j=0}^{n} \frac{(-1)^j \binom{n}{j} (pt)^j N^{(j)}}{N^{(j)}} \cdot [(1+t)^{N-n}(1+t)^{n-j}]$$

$$= (1+t)^{N-n} \sum_{j=0}^{n} (-1)^j \binom{n}{j} (pt)^j (1+t)^{n-j}$$

$$= (1+t)^{N-n}(1+t-pt)^{n},$$

in agreement with (14).

Assume that (17) holds for $m = 1,2,\ldots,k$. Then, from (12), DA

FIL

7 -